

Assessment Development Cycle

All assessments, regardless of method selected, go through the same five stages of development: planning, developing, critiquing, administering, and revising (Figure 4.3). So far in the text, we have discussed three of the four steps in the first stage: determine the intended uses of an assessment, identify the learning targets to be assessed, and select

Figure 4.3 Stages in Assessment Development

1. **Plan:** Assess why? Assess what? Assess how? How Important?
2. **Develop:** Determine the sample. Select, create, or modify test items or tasks and scoring mechanisms.
3. **Critique:** Evaluate for quality.
4. **Administer:** Administer the test or assessment.
5. **Revise:** Evaluate test quality based on results and revise as needed.

the proper assessment method. The fourth step, which we address later in this section, is to determine the relative importance of each learning target so that we sample each adequately.

In the second stage we select or create test items or tasks and scoring mechanisms, adhering to the guidelines offered for each method in Chapters 5 through 8.

During the third stage, we check to make sure we have avoided all possible things that might inadvertently cause results to misrepresent student learning, again using information provided for each method in Chapters 5 through 8.

In the fourth stage, we simply administer the assessment to students.

In the fifth and last stage, we note any problems with the questions, tasks, or scoring mechanisms on the assessment and rework them as needed.

The five stages of development we describe here are presented in the context of a teacher-developed assessment for classroom use. However, they also apply to any other type of assessment developed by grade level teams, content area departments, or district subject-area teams for purposes other than individual classroom use. Short-cycle, common, or interim assessments also need to adhere to standards of quality, and the five stages of development should frame that assessment development process, as well. In Chapters 5 through 8 we will describe any variations on the theme applicable for particular assessment methods.

Stage 1: Plan the Assessment

Creating or selecting a test without having a test plan can result in mismatches between instruction and assessment. The assessment probably will not measure what you intend it to measure, which is known as a *validity* problem. From an assessment quality point of view, this is a bad thing. If you have ever faced an exam yourself that did not match what you thought were the most important aspects of the course you were taking, you know what that feels like from the student's point of view. In the following activity, you will analyze a test you have given to determine its match to the intended learning targets.

When we make a plan for an assessment, whether we intend to create the assessment or just copy it, we are making the advance decisions about validity—what the test will cover and how much weight each learning target will get.

TRY THIS

Activity 4.4 Analyze Your Own Assessment for Clear Targets

Find a selected response test or an extended written response test you have given to students in the past or one you plan to give. Then follow these steps to audit it for clear targets.

1. *Analyze your test item by item.* Identify and write down what learning each item assesses. Describe the learning in whatever terms you want. If two or more items address the same learning, use the same terms to describe that learning. Note the number of test points each item is worth.
2. *Organize the learning targets into a test plan.* Transfer the item information to the chart in Table 4.2 (a printable copy of this table appears on the accompanying CD in the file, “Analyze for Clear Targets”).
3. *Question your test plan.* Is this a representative sample of what you taught and what you expected students to learn?
 - Does the number of points for each learning target represent its relative importance within the whole? If not, which ones are out of balance?
 - Does the number of points for each learning target represent the amount of time you spent on it relative to the whole? If not, which ones are out of balance?
 - Are any important learning targets you taught left out? If so, which one(s)?

Activity 4.4 (Continued)

4. *Adjust your test plan.* As needed, adjust the numbers in the “Number of Questions” and/or “Points” column on the table to reflect the amount of time you spent teaching each learning target and each target’s relative importance to the content as a whole.

As needed, add or delete learning targets to reflect what you taught and what you deemed most important to learn and assess.

5. *Draw conclusions about your assessment.* What does the data you wrote into Table 4.2 tell you about the matches among what's written in your curriculum, what you taught, and what you assessed?

Table 4.2 Analyze an Assessment for Targets

[illegible]

Examples at the Assessment Planning Stage

We'll examine the planning stage by following two examples: a secondary music teacher planning an assessment for his unit on the bluegrass music and a fourth-grade teacher planning an assessment for a unit on the physics of sound. The music teacher's assessment plan is presented in Table 4.3, and the fourth-grade teacher's assessment plan is presented in Table 4.4. As you read these tables, please note that there is no single correct format for test plans; we simply have shown two possible formats.

Table 4.3 Test Plan—Bluegrass Music

Content	Know	Analyze	Compare	Total
Individual pieces of music	5		5	10
Musical elements/instruments/ expressive devices	5	5		10
Composers/performers	5	5	5	15
TOTAL	15	10	10	35

Assess Why?

As we saw in Chapter 2, assessment results can be used for many purposes. In each of our two examples, the teachers' primary purposes are twofold: to help students understand how much they have learned, and to add information to the gradebook in preparation for calculating a course grade. Because assessment design is influenced by how we intend to use the results and by whom else will also use them, we answer the question, "Assess why?" first of all.

Assess What?

Sound assessments arise from clear, specific, and appropriate achievement targets. Beginning with clear targets is important because different targets require different assessment methods and also because the breadth and depth of a learning target will affect how much coverage it will need on the assessment and in instruction. So at this juncture, you will do the following:

1. List the major learning targets you will be teaching.
2. Identify the prerequisite subtargets by unpacking or clarifying the learning targets, as needed.
3. Classify the targets, subtopics, and/or unpacked learning targets, into knowledge, reasoning, performance skills, products, and/or dispositions.
4. Write the unpacked and/or clarified learning targets into the appropriate spaces in the test plan format you select. Blank forms are on the CD in the file, "Test Planning Forms."

Table 4.4 Test Plan—The Physics of Sound, for Fourth Graders (Selected Targets)

Learning Target	Type of Target	Assessment Method	Percent Importance
Acquire vocabulary associated with the physics of sound	Knowledge	Selected Response	25%
Learn that sound originates from a source that is vibrating and is detected at a receiver such as the human ear	Knowledge	Selected Response	5%
Use knowledge of the physics of sound to solve simple sound challenges	Reasoning	Extended Written Response	20%—Present novel sound challenge; student describes how to solve it
Understand the relationship between the pitch of a sound and the physical properties of the sound source (i.e., length of vibrating object, frequency of vibrations, and tension of vibrating string)	Reasoning	Extended Written Response	10%—Give two novel examples and student compares pitch
Use scientific thinking processes to conduct investigations and build explanations: observing, comparing, and organizing (1) How sound travels through solids, liquids, and air; (2) Methods to amplify sound at the source and at the receiver	Reasoning Skill	Extended Written Response & Performance Assessment	40%—Design an experiment for a given hypothesis; give data/student organizes; set up stations/students conduct an experiment—all novel

Source: From the *FOSSR Physics of Sound Teacher Guide*, © The Regents of the University of California, 2005, developed by Lawrence Hall of Science and published by Delta Education, LLC. Reprinted by permission.

The secondary school music teacher whose test plan is represented in Table 4.3, has planned a 3-week unit of instruction on bluegrass music. He has chosen bluegrass music as the context for the following music standards:

- Classifies selected exemplary works from various historical periods by genre, style, and composer.
- Explains how use of specific musical elements (for example, rhythm, melody, timbre, expressive devices) is characteristic of music from various world cultures.
- Identifies music that represents the history and diverse cultures of our state.
- Identifies important composers and performers who influenced various genres of American music.

Students will need to acquire some knowledge about bluegrass music in three categories—works (famous pieces of music), musical elements (used to give the music the bluegrass feel), and composers/performers. In addition, the teacher will teach students to use the content knowledge in each of these three areas to reason analytically and comparatively. As indicated in the test plan, any single test question either will test knowledge or will be a combination of knowledge and the reasoning that is to be performed using that knowledge.

In the plan for the fourth-grade unit on the physics of sound, the teacher has written selected learning targets down the left-hand column of Table 4.4. The type of learning target is noted in the next column.

These teachers chose content categories based on their state content standards, local curriculum guides, and natural subdivisions of content. They chose reasoning patterns from content standards, local curriculum, and priorities in their teaching.

Assess How?

This is fairly straightforward. Once you have classified learning targets by type it is easy to decide which assessment method to select by referring to the matching guidelines in Table 4.1.

The fourth-grade teacher is emphasizing science process skills as well as knowledge and reasoning so she will be using more than one assessment method. She has chosen the planning format shown in Table 4.4, which allows her to specify how each learning target will be assessed. The music teacher has only knowledge and reasoning learning targets. He has decided that the combination of knowledge and reasoning can be assessed well with a

selected response test. Since he has no need for a test plan to show different assessment methods, he has chosen a test plan format that emphasizes how content knowledge crosses with level of thinking.

How Important?

When we define the relative importance of each of the learning targets listed, we are mapping out how we will *sample* student learning. What will be most important on this assessment? How many points will each item be worth? For the most part, this is the call of the individual teacher, taking into account the following:

- *The breadth and depth of the learning target.* For example, in Table 4.4, the learning target “Learn that sound originates from a source that is vibrating and is detected at a receiver such as the human ear” doesn’t cover as much territory as “Acquire vocabulary associated with the physics of sound,” or “Use scientific thinking processes to conduct investigations and build explanations: observing, comparing, and organizing.” Therefore, assessing “learning where sound originates” will carry less weight on the assessment, as reflected by the percentage of total points, and other targets will carry more weight.

In all cases, the assessment must include enough questions or tasks to provide evidence leading us to a confident conclusion about student achievement, without wasting time gathering too much evidence. The critical question is, How much evidence is enough? How many multiple-choice test items, essay exercises, performance tasks? (Each assessment method brings with it a set of rules of evidence for determining how big a sample of student achievement we need. We explain those guidelines in Chapters 5 through 8.)

- *The importance of each learning target.* For example, in Table 4.4, the teacher has determined that the most important learning target focuses on science processes and skills. Scientific information is important, and there is an expectation that students will learn some content information from this unit of study, but process skills are more important in this case. Therefore, science process targets alone will comprise 40 percent of the assessment points and the other four targets combined will total 60 percent.
- *State standards and local curriculum.* For example, the music teacher is guided by the state standard in his emphasis of knowledge and reasoning targets in the unit. Because the state standards emphasize using information to analyze and classify, the teacher has also emphasized it on his test—two-thirds of the points on the test reflect students’ ability to apply knowledge in novel ways.

Although not a hard and fast rule, a good guideline for making decisions regarding percentage of importance for each learning target is that percentage of instructional time and percentage of assessment time should be roughly equal. So, if science processes and skills represent 40 percent of importance, roughly 40 percent of instructional time will be used to teach science processes and skills.

Stage 2: Develop the Assessment—Good Exercises and Sound Scoring Procedures

Having chosen one or more assessment methods, we must adhere to guidelines for developing test questions, extended written response exercises, performance tasks, or questions to elicit evidence of the desired student learning. Further, we need accurate scoring keys for selected response assessments and good scoring guides for extended written response and performance assessment. The development guidelines for each are covered in detail in Chapters 5 through 8.

Stage 3: Critique the Assessment

We've carefully selected and clarified learning targets, determined their relative importance, chosen the best assessment method, and carefully written questions, tasks, and scoring guides adhering to standards of quality. We're finished, right?

Well, no. Regardless of how carefully we plan, things can still go wrong that result in inaccurate estimates of achievement. Witness the pumice that wouldn't float in one state's performance assessment of science process skills, the writing prompt that elicited expository instead of persuasive writing from students, the 10th-grade standardized test administered during rioting by seniors, or asking English language learners to write extended responses to math problems.

A list of problems that can result in inaccurate estimates of student learning is presented in Figure 4.4. Note that some of the problems listed, such as unclear targets, inappropriate assessment method, and improper sampling, would be solved by adhering to the test development process as discussed here. Others are crazy little things that can be hard to anticipate if you haven't experienced them. Problems specific to each method will be discussed more fully in Chapters 5 through 8.